FBS25- Identifiler® Plus Data Analysis Using STRmix™

Table of Contents

1. Scope

1.1. This method describes the process by which Identifiler® Plus data generated from Applied Biosystems GeneMapper® ID-X Software (GMID-X) is entered into STRmix™.

2. Background

2.1. STRmix™ is a probabilistic genotyping system based on a biological model, statistical theory and computer algorithms. Probabilistic genotyping is a tool used to assist the DNA analyst in the interpretation of DNA typing results. STRmix™ uses a fully continuous approach to interpret DNA profiles including mixture deconvolution. The software can compare reference DNA profiles to casework profiles to generate a measure of weight of the evidence (likelihood ratio) in relation to a pair of propositions or it can perform mixture deconvolutions if there are no reference DNA profiles to compare.

3. Safety

3.1. Not applicable

4. Materials Required

4.1. GeneMapper® ID-X Software, versions 1.3 or 1.4

4.2. STRmix™ Software, version 2.3

4.3. Windows-based computer capable of running the software

5. Standards and Controls

5.1. Not applicable

6. Procedures

6.1. Software running options

6.1.1. STRmix™ analysis can occur on the analysts' personal desktop computer or by connecting to the server.

6.1.1.1. For single source and simple 2-person mixtures, run STRmix™ on personal desktop computer.

6.1.1.1.1. **Note:** The maximum number of MCMC chains which can be run should match the maximum number of available cores on the computer.

6.1.1.2. For complex mixtures or samples which will not run on the personal desktop computer, run STRmix™ by connecting to the server.

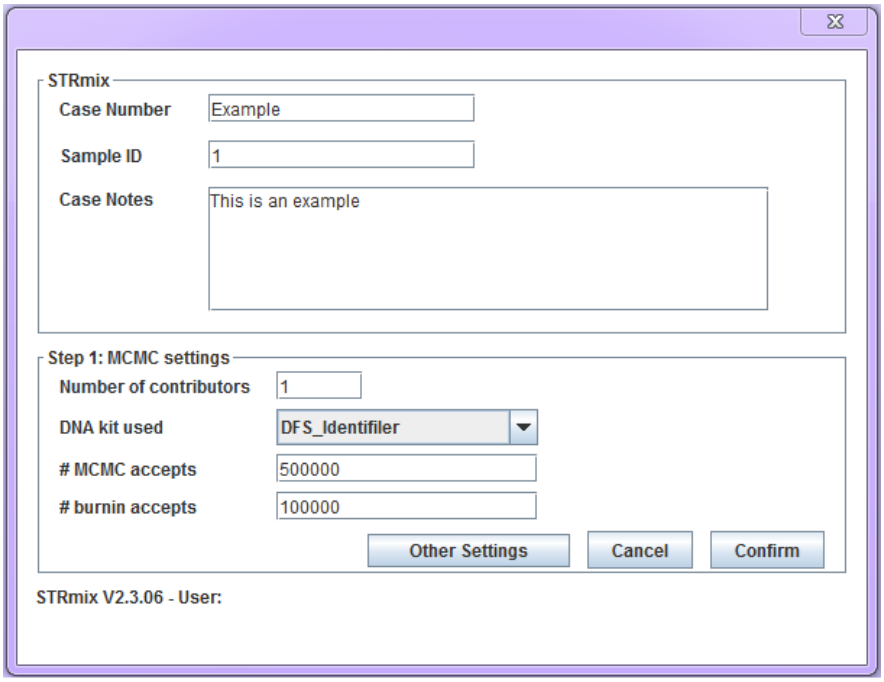
6.1.1.2.1. **Note:** The maximum number of chains which can be run on the server is 8. The analysts may increase this number from 4 to 8 in multiples of 2 to decrease run time. Refer to section 6.3.5.

6.2. Launching STRmix™

6.2.1. Open the STRmix™ software by locating STRmix™ in the task bar or by double clicking on the STRmix™ icon located on the desktop. Either action will display the STRmix™ main menu (see diagram below).

6.3. Start Mixture Analysis

6.3.1. Select **Start Mixture Analysis** to open the Sample Summary window (see diagram below).



6.3.2. Complete Case Number and Sample ID information in the case notes section. **Note:** Additional case notes may be entered into the Case Notes section.

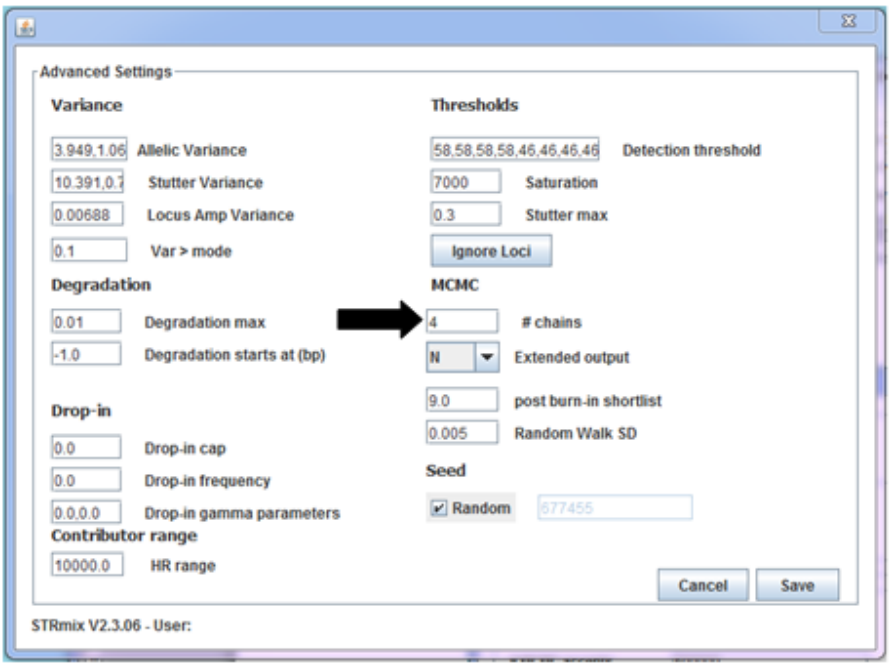
6.3.3. Enter Number of contributors and confirm that the correct DNA kit (i.e. DFS_Identifiler) is selected. **Note:** See FBS21 Identifiler® Plus ID Interpretation to determine the number of contributors.

6.3.4. Verify that the number of MCMC accepts = 500000 and the number of burnin accepts = 100000.

6.3.4.1. **Note:** For single source samples with good peak heights and where variability is expected to be extremely low, the analyst may lower the number of MCMC and Burnin accepts (iterations) to 50,000 and 10,000, respectively.

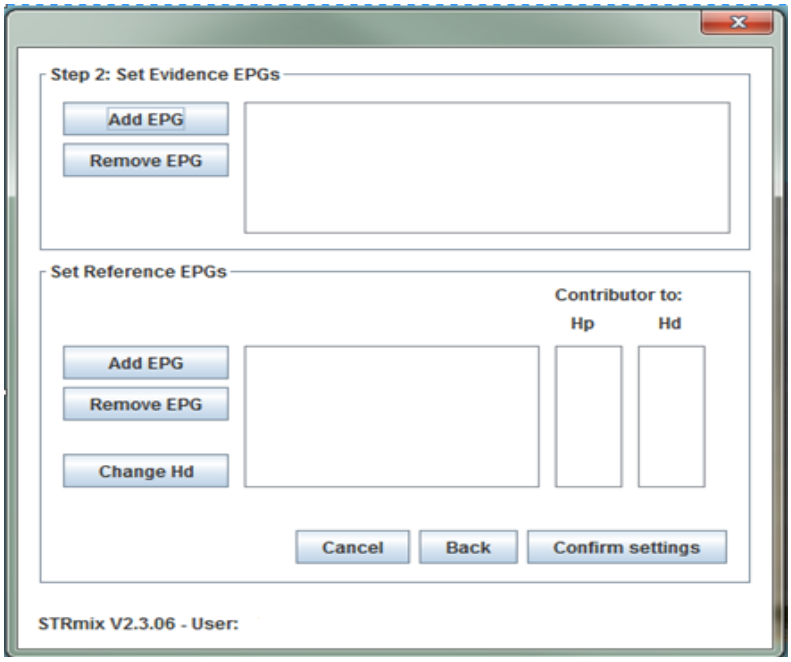
6.3.4.2. **Note:** To increase the value of MCMC accepts and Burnin accepts see section 6.9.2.

6.3.5. **Note:** For samples that are run on the server (see section 6.1.1.2), verify that the number of MCMC chains is 8 by selecting **Other Settings** and then Advanced Settings. Proceed with the subsequent steps.

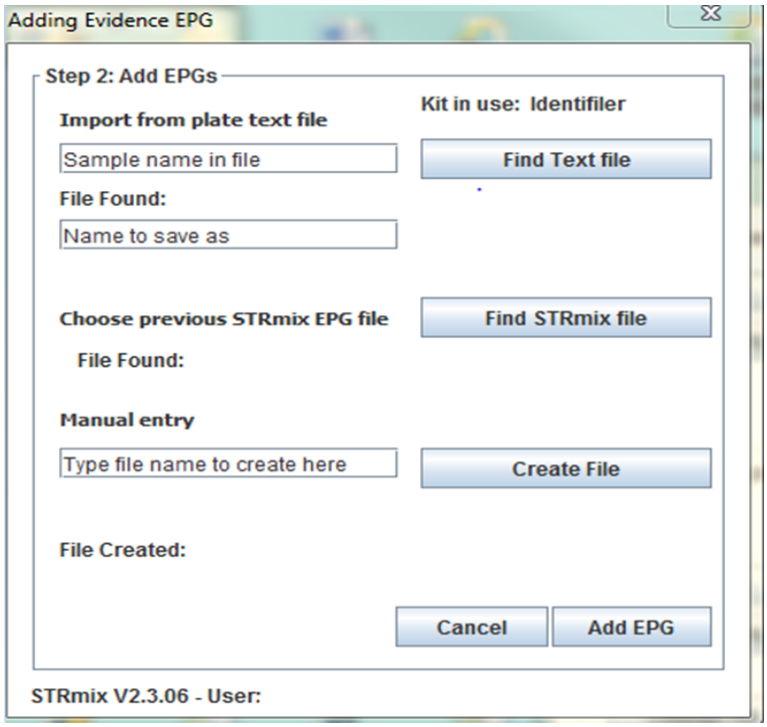


6.3.6. Select **Confirm** to proceed to the Set Evidence/Set Reference EPGs window or **Cancel** to return to the Startup screen.

6.3.7. In the Set Evidence/Set Reference EPGs window (see diagram below) select **Add EPG** from the Set Evidence EPGs section.



6.3.8. In the Adding Evidence EPG window (see diagram below), select **Find Text file** to enter an evidence input file.



6.3.8.1. For a single source sample with an associated reference sample, conducting a deconvolution is not necessary. Proceed to step 6.3.9 to add the reference sample.

6.3.8.2. For mixtures with an associated reference sample(s) and no assumption of contributor(s) (i.e. no conditioning), conduct a deconvolution (no calculation of an LR) prior to entering any reference(s) into STRmix™. Click **Confirm Settings** on the Set Evidence/Set Reference EPGs screen and proceed to step 6.3.11.1.

6.3.8.3. For mixtures with an associated reference sample(s) in which assumed contributor(s) will be conditioned upon, click **Confirm Settings** and proceed to step 6.3.9 to add the assumed contributor reference sample(s).

6.3.8.4. If conducting a deconvolution on mixtures without an associated reference sample(s) click **Confirm Settings** on the Set Evidence/Set Reference EPGs screen then proceed to step 6.3.11.1.

6.3.9. To add a reference sample select **Add EPG** from the Set Reference EPGs section of the Set Evidence/Set Reference EPGs window.

Note: If using a reference sample typed with the GlobalFiler amplification kit, follow the instructions on the Conversion Sheet (Document Control Number: 6780) prior to adding the reference EPG to STRmix. Use of the Conversion Sheet is acceptable for comparison references only. Converted reference samples may <u>not</u> be conditioned on as assumed contributors in the STRmix analysis.

6.3.10. In the Adding Evidence EPG window, select **Find Text file** to enter a reference input file.

6.3.10.1. If conducting an LR for a single source sample proceed to step 6.3.11.

6.3.10.2. If conducting a deconvolution with assumed contributor reference(s) proceed to step 6.3.10.5.

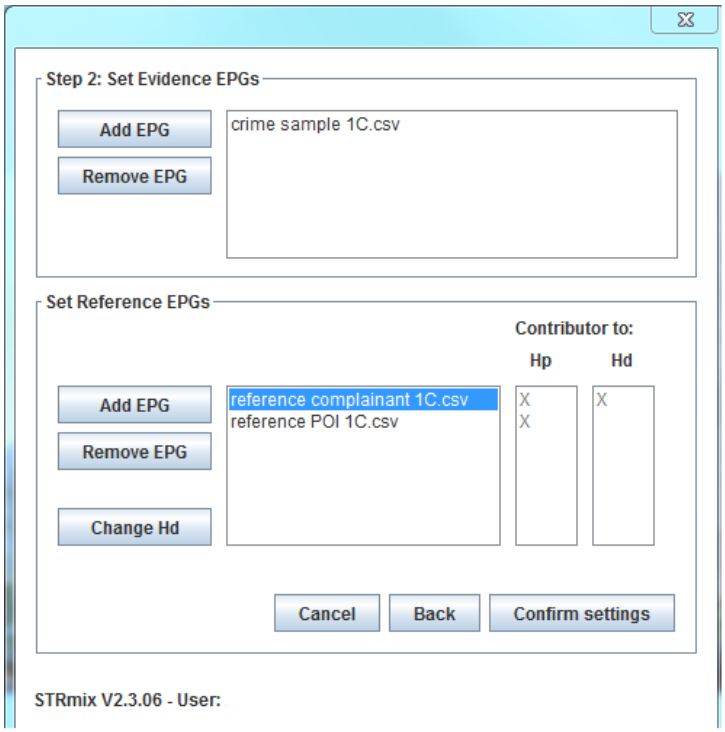
6.3.10.3. If conducting an LR from Previous proceed with the next steps.

6.3.10.3.1. The entering of multiple Persons of Interest (POI's) into STRmix™ will be case dependent. For example, when two or more POI's are included in a mixture, typically the POI's would be run through STRmix™ together, then each run individually. For instances when multiple scenarios are run through STRmix™, include all advanced reports in the case file and

report the scenario that most accurately describes the results.

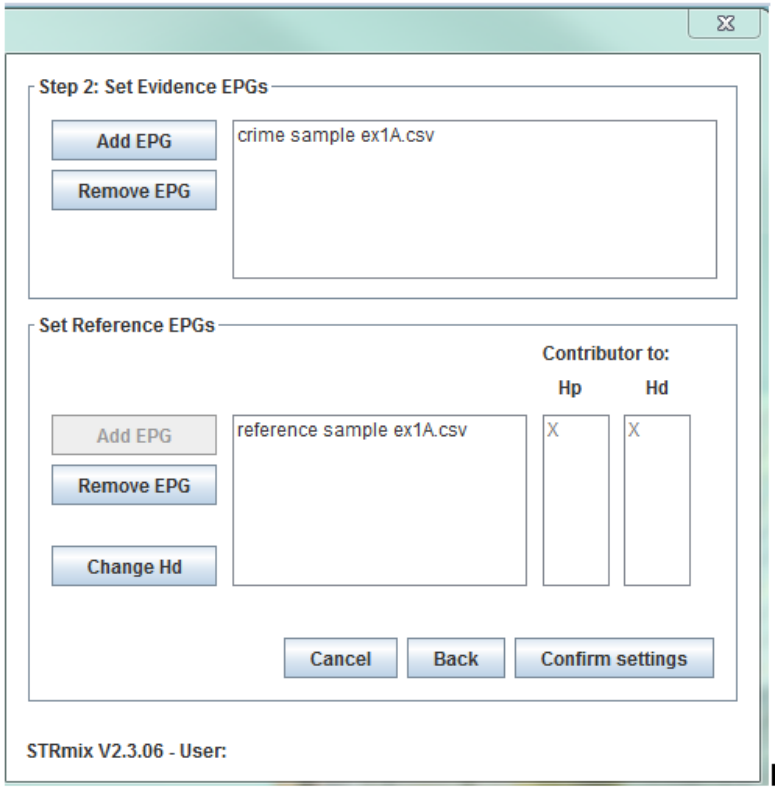
6.3.10.3.1.1. Repeat steps 6.3.9 and 6.3.10 to add additional references.

6.3.10.3.2. If there are multiple references in which assumed contributor(s) will be conditioned upon, the assumed contributor(s) will be input first. Condition the assumed contributor(s) by highlighting the input file then selecting **Change Hd** so that both Hp and Hd are marked with an X (see diagram below). Select **Confirm settings**.

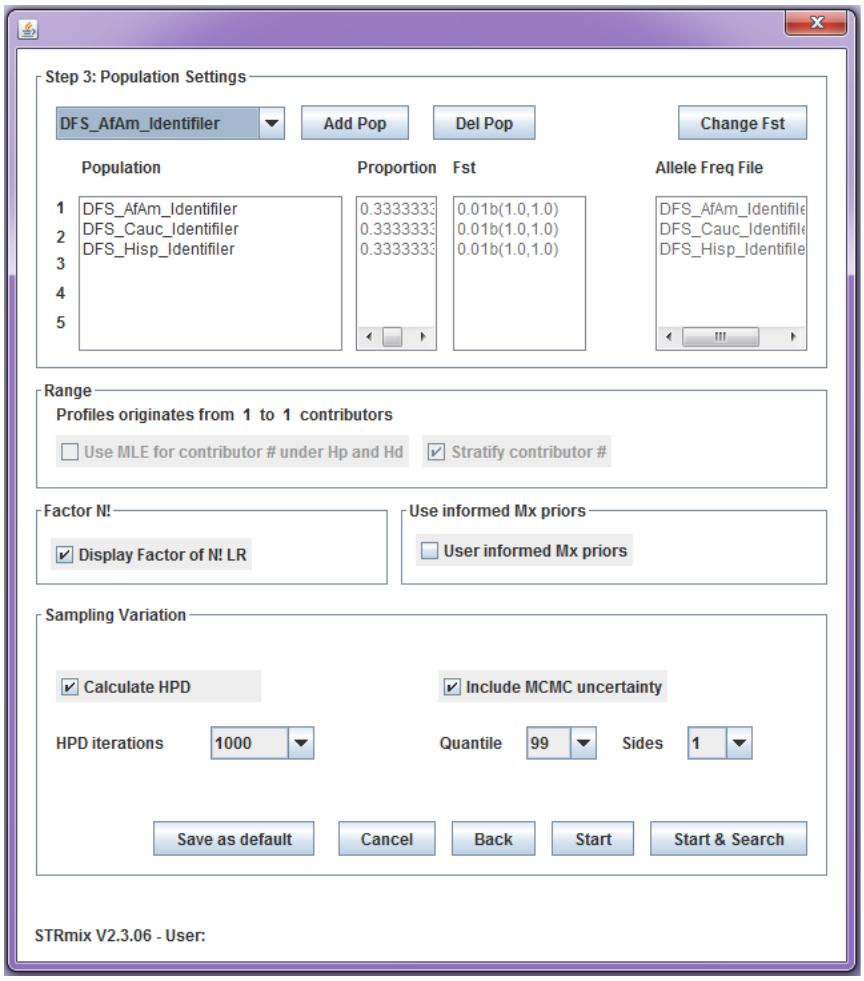


6.3.10.4. If no conditioning, ensure that the Hd column is not marked with an X (see diagram below).

6.3.10.5. When conducting a mixture deconvolution (no calculation of an LR) and there is an assumed contributor reference sample, condition the assumed contributor by highlighting the input file then selection **Change Hd** so that both Hp and Hd are marked with an X (see diagram below). Select **Confirm settings**.



6.3.11. From the Population Settings window (see diagram below), select the appropriate populations from the drop down menu.

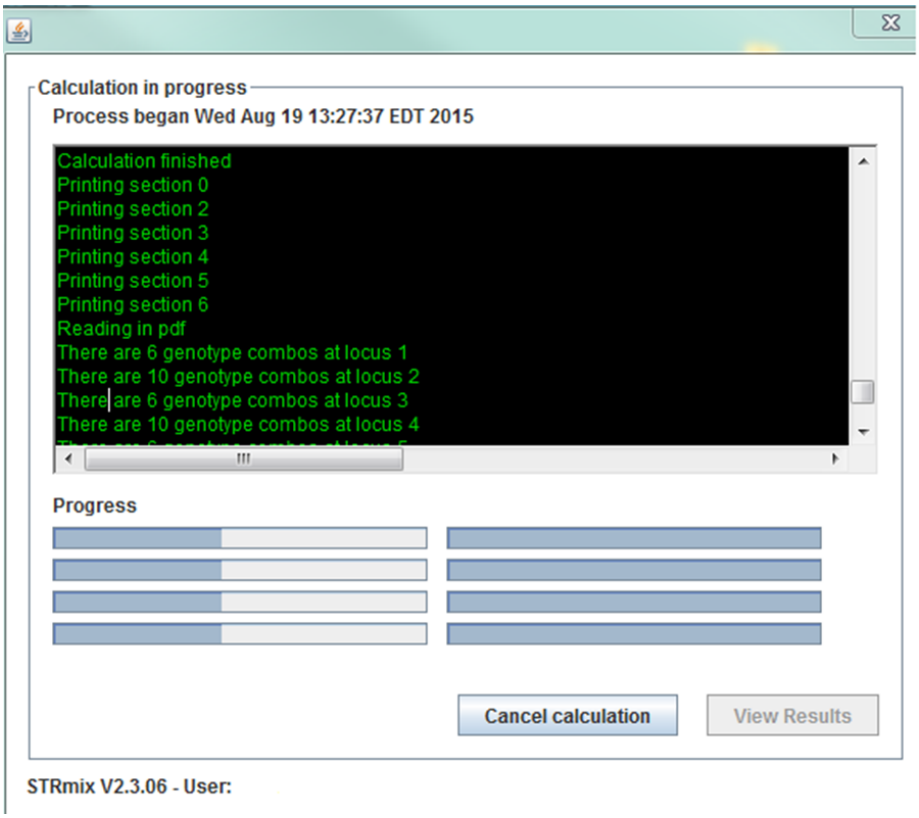


6.3.11.1. If performing a mixture deconvolution without a reference standard the Population Settings are inaccessible (grayed out). Select **Start** to continue and proceed to step 6.3.14.

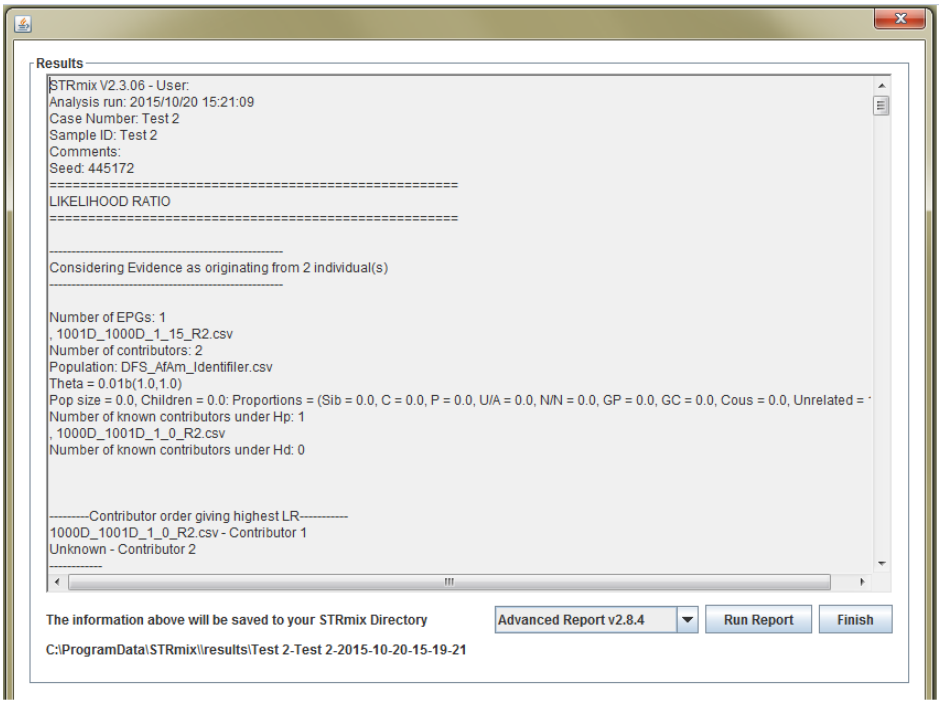
6.3.12. **Note:** With the exception of the User informed Mx priors option, all other settings on the Population Settings screen are default and shall not change. The User informed Mx priors option can only be selected with approval from the Technical Leader.

6.3.12.1. The three default populations for the DFS_Identifiler are DFS_AfAm_Identifiler, DFS_Cauc_Identifiler, and DFS_Hisp_Identifiler.

6.3.13. Select **Start** and the Calculation in progress screen will appear (see diagram below). **Note:** In the Population Settings window, you may select **Cancel** to return to the Startup screen or **Back** to go to the previous window.



6.3.14. Upon completion of the calculation, a summary of the analysis results will be generated and appear automatically (see diagram below).



6.4. Advanced Report

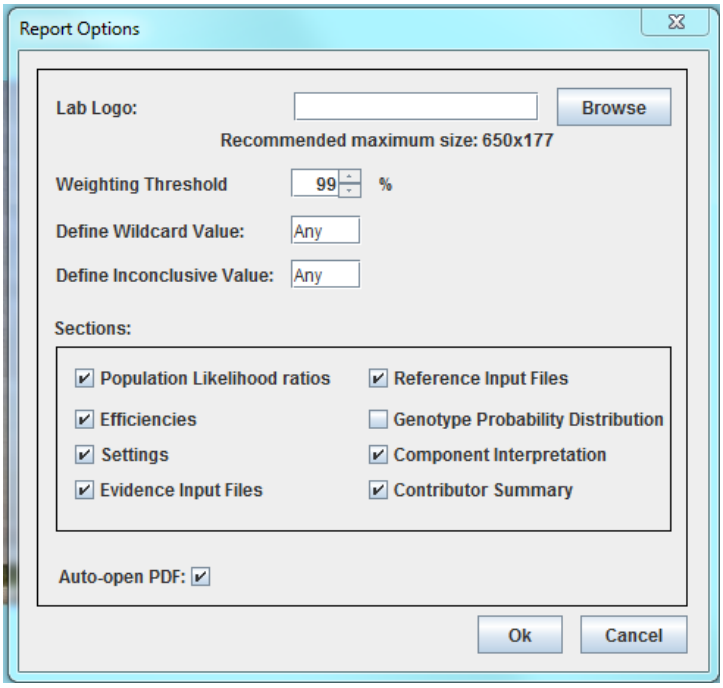
6.4.1. Select **Run Report** to view the Advanced Report.

6.4.2. Once the Advanced Report window opens, Select **Options** to ensure the Report Options default settings are correct (see diagram below).

6.4.3. Displayed below are the Report Options default settings.

6.4.3.1. **Note:** If using the server, the Report Options default settings are the same as below; however, the "Auto-open PDF" option should be unselected. Ensure the "Auto-open PDF" option is unselected.

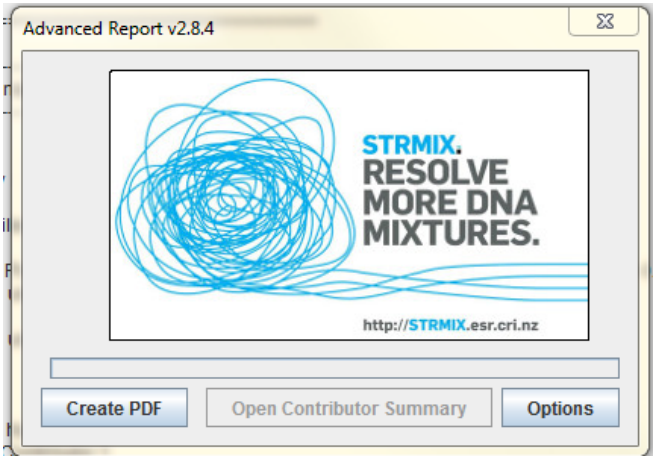
6.4.3.2. **Note:** Do not alter the default settings.



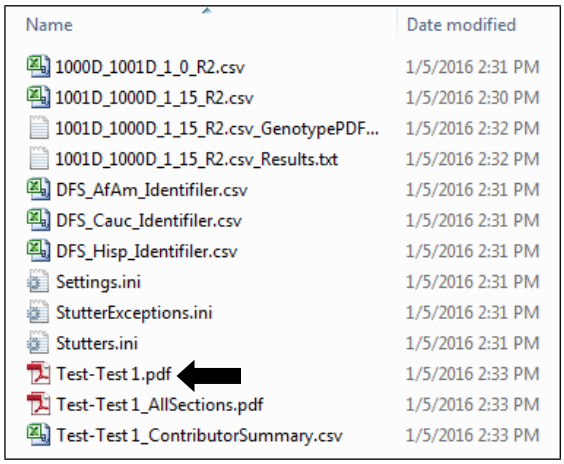
Note: The designation of "Any" in the Component Interpretation section of the Advanced Report indicates that the weighting threshold has not been met during the deconvolution process.

6.4.4. Select **Ok** to return to the Advanced Report window.

6.4.5. Once the Advanced Report window opens, select **Create PDF** (see diagram below) and save the report to an appropriate location.

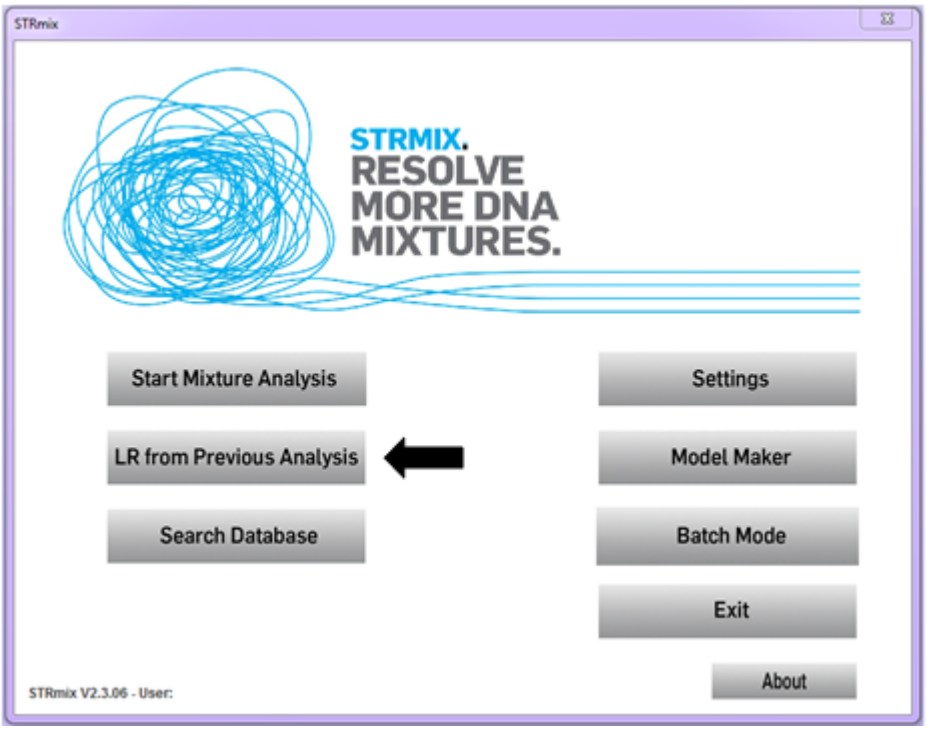


6.4.6. Print the PDF Advanced Report. **Note:** Two versions of the Advanced Report will be saved to the appropriate location. Print the **non**-"AllSections.pdf" version (see diagram below).

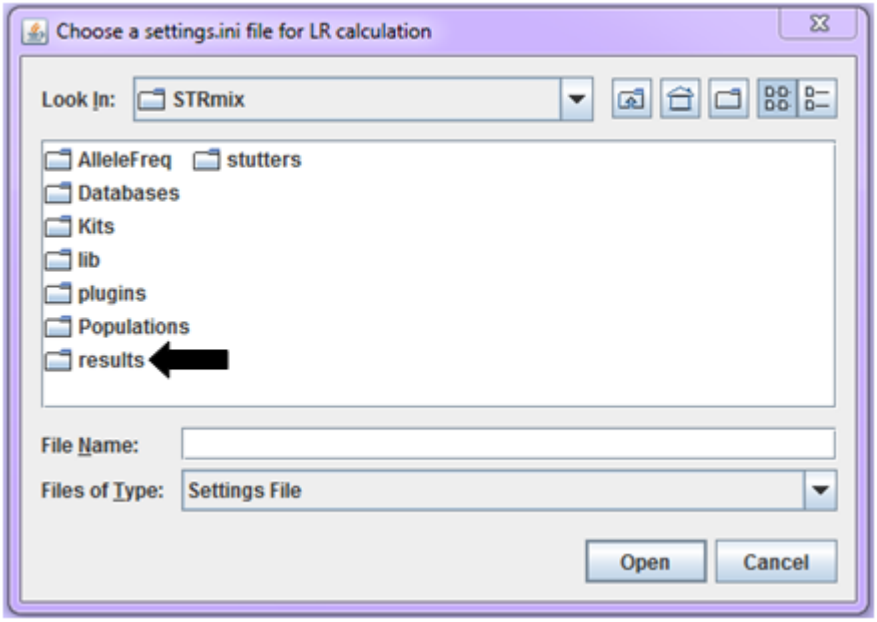


6.5. LR from Previous Deconvolutions

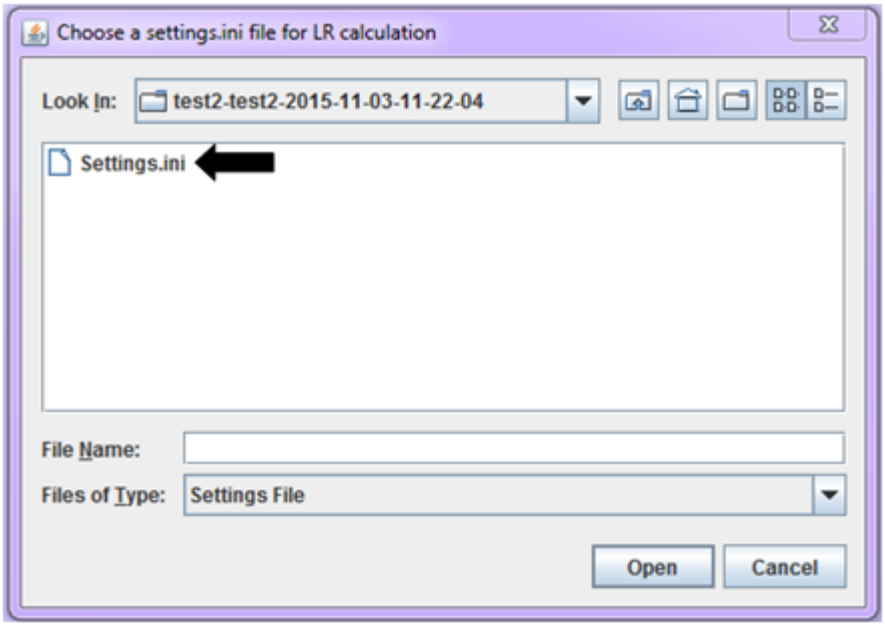
6.5.1. To perform an LR from previous for mixture deconvolutions, select **LR from Previous Analysis** from the STRmix™ main menu (see diagram below).



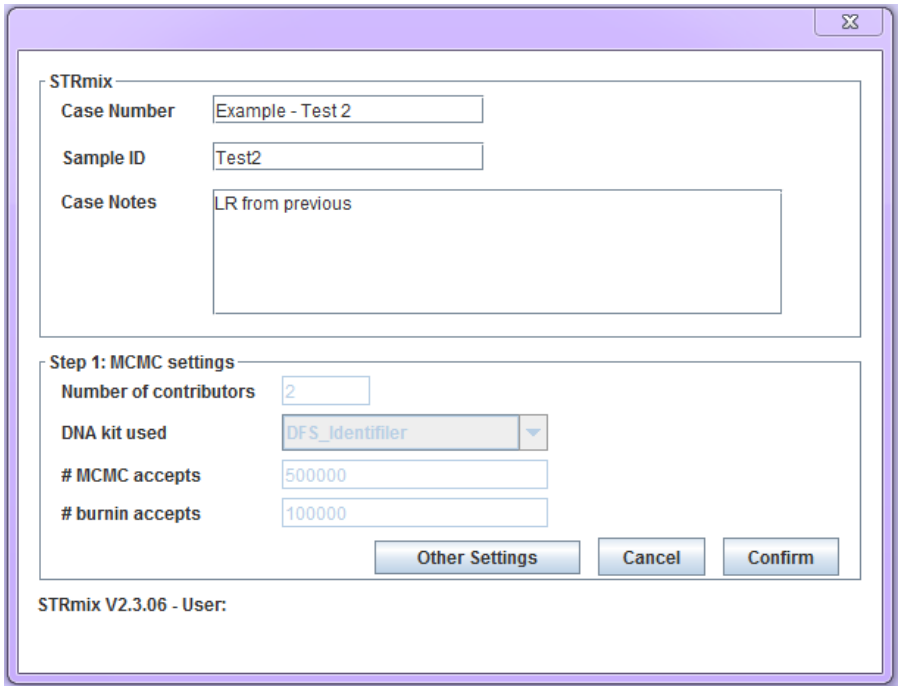
6.5.2. Open the Results folder and navigate to the appropriate results folder from the deconvolution of interest (see diagram below).



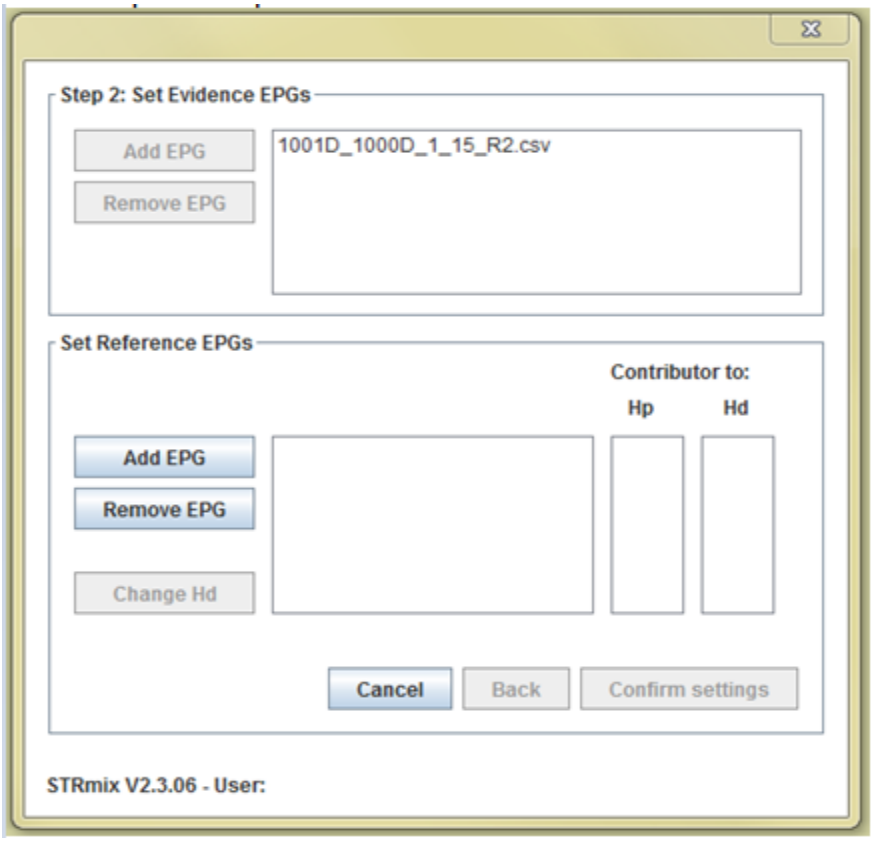
6.5.3. Open the Settings.ini file (see diagram below).



6.5.4. Ensure that the correct file was opened then select **Confirm** in the STRmix™ window to proceed (see diagram below). In the Case Notes section, enter "LR from previous" or add "LR from previous" to the Sample ID. **Note:** Additional case notes may be entered into the Case Notes section.



6.5.5. Add references. **Note:** The evidence EPG sections will be inaccessible (grayed out) (see diagram below). See section 6.3.9 to add references then proceed with the subsequent steps.



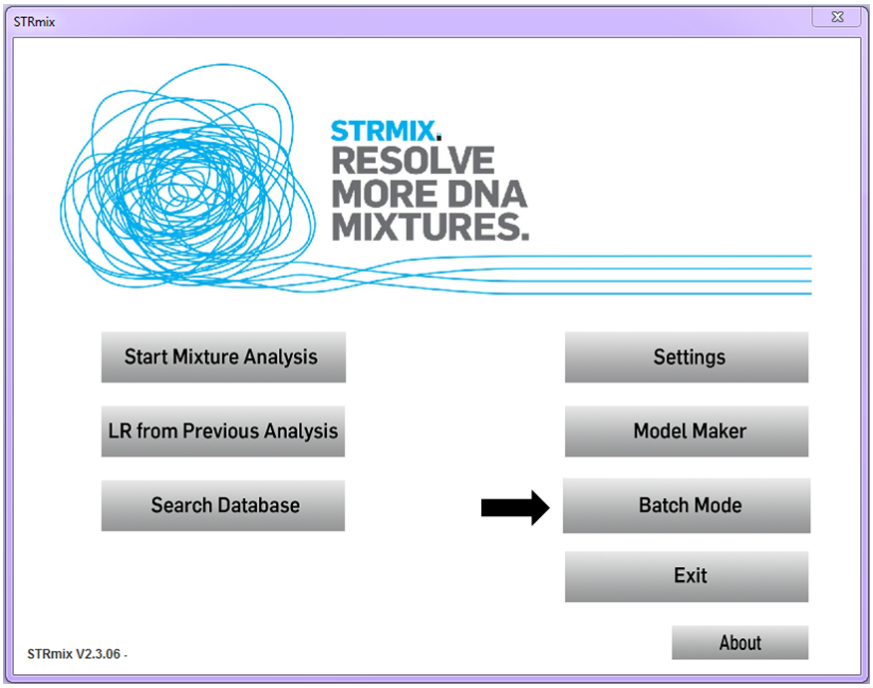
6.6. Multiple STRmix™ analyses

6.6.1. Each time a profile is run through STRmix™ analysis the results will vary slightly. In order to be as unbiased as possible STRmix™ analysis of a single profile should only be conducted once, with those results being reported unless troubleshooting is needed.

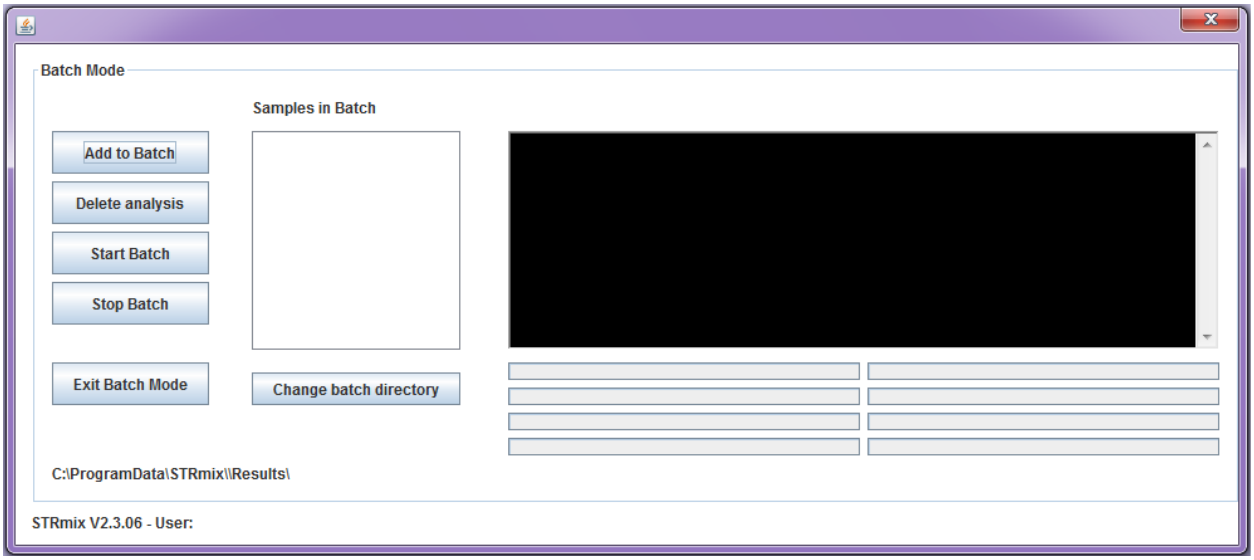
6.7. Batch Mode STRmix™ analyses

6.7.1. A number of STRmix™ analyses can be set up and queued to run sequentially.

6.7.1.1. To set up a queued analysis for multiple samples, select **Batch Mode** from the STRmix™ main window (see diagram below).



6.7.1.2. Select **Add to Batch** from the Batch Mode window (see diagram below) to open the Sample Summary window.



6.7.1.3. Complete the analysis set up for the first sample following steps 6.3.2 through 6.3.12.

6.7.1.4. In the Population Settings window, select **Start** to return to the Batch Mode window.

6.7.1.5. In the Batch Mode window, select **Add to Batch** to enter the next sample. Repeat steps 6.7.1.3 through 6.7.1.4 to add additional samples.

Note: To remove a sample from the batch mode, highlight the case/sample in the Samples in Batch section of the Batch Mode window then select **Delete analysis**.

6.7.1.6. Select **Start Batch** to start the batch run.

6.7.1.7. After completion of analyses, select **Exit Batch Mode** to return to the STRmix™ main window.

Note: Each of the analyses details will be saved to the designated default location.

6.8. Reviewing STRmix™ data

6.8.1. In the Advanced Report, reviewers should check:

6.8.1.1. Summary of Input Data:

6.8.1.1.1. The correct number of contributors has been selected.

6.8.1.1.2. The correct input file(s) has been selected.

6.8.1.2. Parameters:

6.8.1.2.1. The correct settings have been used.

6.8.1.3. Run Information:

6.8.1.3.1. Check the total number of iterations. The value displayed indicates the total number of post-burn in iterations that the MCMC has run during its analysis. This value, along with the number of accepts chosen for analysis can inform the user as to how often a new proposed set of parameters was accepted. This is referred to as the acceptance rate. The acceptance rate is calculated by dividing the number of post burn-in accepts by the total number of iterations.

For example, if total iterations = 4,505,505; burn-in accepts = 100,000; and total accepts = 500,000, the acceptance rate calculation would be as follows:

400,000/4,505,505 = 0.088 or 1 in 11.3.

A very low acceptance rate (e.g. 1 in thousands to millions) may, in combination with the other diagnostics, indicate that the analysis needs to be run for additional iterations.

Note: On its own (and without any other indication of sub-optimal results) a low acceptance rate is not an indication that rework is required.

6.8.1.3.2. Check the effective sample size. Effective sample size (ESS) is the number of independent samples the MCMC has taken from the posterior distribution of all parameters. A low ESS in relation to the total number of iterations suggests that the MCMC has not moved very far with each step or has had a low acceptance rate. A low ESS value (e.g. 10s or 100s) means that there is potential for a large difference in weights if the analysis was run again. This potential will be taken into account during HPD interval generation in any LR calculations (unless the genotype sets are completely resolved on a single combination, in which case there will be no effect of ESS on the HPD interval).

Note: On its own (and without any other indication of sub-optimal results) ESS is not an indication that rework is required.

6.8.1.3.3. Check the average log (likelihood). This value shows the average $\log_{10}$ (likelihood) for the entire post burn-in MCMC. This is the log of the average likelihood (or probability) value created at each of the post burn-in MCMC iterations. The larger this value, the better STRmix™ has been able to describe the observed data. A negative value suggests that STRmix™ has not been able to describe the data very well given the information it has been provided. The following are reasons why this value may be low or negative:

6.8.1.3.3.1. The profile is simply very low level and there is very little data making up the likelihood.

6.8.1.3.3.2. The number of contributors is wrong and there are forced stochastic events in the STRmix™ run as a result (e.g. large heterozygote peak imbalances or variations in mixture proportions across the profile).

6.8.1.3.3.3. Data has been removed that was real, particularly stutter peaks, and must now be described in STRmix™ by dropout.

6.8.1.3.3.4. Artifact peaks have been left labeled and must now be accounted for in STRmix™ by drop-in.

6.8.1.3.3.5. A low or negative value for the average (log_{10}) likelihood may indicate to users that the analysis requires additional scrutiny.

Note: Good quality mixed DNA profiles are likely to give higher average log_{10} (likelihood) values than good quality single source profiles; therefore, low average log_{10} (likelihood) values alone are not necessarily an indicator of an issue.

6.8.1.3.4. Check the Gelman-Rubin convergence diagnostic value. This diagnostic informs the user whether the MCMC analysis has likely converged. If this value is above 1.2 then it is possible that the analysis has not converged. Refer to the troubleshooting section 6.9.2.

6.8.1.3.5. Check the allele variance and stutter variance. Both of these values are the average value for allele variance and stutter variance constants across the entire post burn-in MCMC analysis. These values can be used as a guide as to the level of stochastic variation in peak heights that is present in the profile.

If the variance constant has increased markedly from the mode of the prior distribution, then this may indicate that the DNA profile is sub-optimal or that the number of contributors is incorrect. Refer to the Parameters section of the advanced report to obtain the mode value for comparison.

Used in conjunction with the average $\log_{10}$ (likelihood), a large allele variance or stutter variance constant can indicate a poor PCR.

If the sample is simply low level this should result in a low average $\log_{10}$ (likelihood) and an average variance constant.

If some data has been omitted, left on or misinterpreted this should result in a low average $\log_{10}$ (likelihood) and high variances.

6.8.1.4. Summary of Contributors/Efficiencies/Component Interpretation:

6.8.1.4.1. The mixture proportions, degradation, and LSAE all appear correct when compared to the EPG(s).

6.8.1.5. Component Interpretation:

6.8.1.5.1. The weightings make intuitive sense.

6.8.2. For each LR calculation, reviewers should check:

6.8.2.1. Reference Input File(s):

6.8.2.1.1. The correct reference has been compared.

6.8.2.2. Summary of Input Data/Comments:

6.8.2.2.1. The correct assumptions have been made, if applicable.

6.8.2.2.2. The correct hypotheses have been used.

6.8.2.3. Summary of LR/Per Locus Likelihood Ratios:

6.8.2.3.1. The LR broadly agrees with a human interpretation of assessing the potential contribution of an individual.

6.9. Troubleshooting

6.9.1. It is important for STRmix™ analysis results to be checked by examining the weightings of various genotypes and the DNA profile(s) observed. There may be instances when the results obtained do not seem intuitively correct.

6.9.1.1. The following are examples of this:

6.9.1.1.1. Large LR's (>1) are obtained for each locus, except one where the LR = 0 and the POI reference is consistent with the evidentiary profile.

6.9.1.1.2. The mixture proportions do not reflect what is observed.

6.9.1.1.3. The degradation does not reflect what is observed.

6.9.1.1.4. The interpreted contributor genotypes are not intuitively correct.

6.9.1.2. Causes for the above examples may be due to the following:

6.9.1.2.1. The MCMC has not run for enough iterations.

6.9.1.2.2. The number of contributors has not been correctly chosen.

6.9.1.2.3. The PCR process has been affected (i.e. inhibition).

6.9.1.3. Should the weights and/or diagnostics imply to the analyst that further scrutiny is required then a number of re-work options are available. For example, a review of the proposed number of contributors should be considered. Further analytical work such as a re-amplification to strengthen the number of contributors assumption or to assist with allele designation/sub-optimal PCR performance. An analyst may also increase the total number of iterations if the acceptance rate is low (section 6.8.1.3.1), the ESS is low (section 6.8.1.3.2), and/or the Gelman-Rubin value is significantly above 1.2 (sections 6.8.1.3.4 and 6.9.2).

6.9.2. The Gelman-Rubin Convergence Diagnostic value needs to be checked. If this value is above 1.2, STRmix™ analysis may be repeated under the same conditions or run with an increased number of MCMC and Burnin accepts (iterations) e.g. 5,000,000 and 1,000,000, respectively. Consult with Technical Leader prior to increasing number of iterations.

6.9.3. Instances may occur when the complexity of the DNA mixture being analyzed cause the computer to run slow or stop. If the run fails or a memory error occurs, the profile may need to be run on the server. If the server continues to run slow or has stopped, carefully evaluate the profile to confirm that all artifacts have been removed, an appropriate number of contributors has been estimated, whether all loci should be considered and/or whether the entire profile should be considered uninterpretable.

6.9.4. If an error message is obtained from the software, it can be related to errors in the input file, lack of computer memory, or user input error during setup. Each scenario should be investigated to determine cause and make necessary adjustments. See the STRmix™ internal validation report and/or STRmix™ User's Manual for more specific descriptions of the errors observed and their determined causes.

7. Sampling

7.1. Not applicable

8. Calculations

8.1. Not applicable

9. Uncertainty of Measurement

9.1. Sampling uncertainty occurs due to the finite allele probabilities that are associated with the population samples being used. In STRmix™ an allowance for sampling uncertainty is implemented by adjusting the allele frequencies using a Bayesian posterior mean frequency (i.e. Highest posterior density (HPD)) to better account for sampling uncertainty with allele counts within a limited population data set.

9.2. The combined LR calculated by STRmix™ is referred to as a point estimate. Because the true answer is not known, a credible interval is then applied around the point estimate known as a one-sided 99% HPD credible interval. This interval accounts for the uncertainty associated with the point estimate LR. This interval, commonly applied in Bayesian statistical calculations, give a range (i.e., with 99% credibility) of where the true allele proportions actually lie. The lower end of the HPD interval is reported from STRmix™ to be most conservative to the accused or person of interest.

9.3. STRmix™ uses the Balding and Nichols model (NRC II recommendation 4.2) to account for uncertainty with alleles for a genotype at a locus that may be identical by decent. The use of this model allows for correction of sub-populations effects.

10. Limitations

10.1. STRmix™ cannot incorporate mutations such as primer binding site mutations, trisomies or somatic mutation. When these effects are present, the locus must be ignored during the profile analysis in STRmix™.

10.2. STRmix™ cannot incorporate multiple EPGs generated via different STR kits into one analysis.

10.3. An artifact peak in the EPG will be considered as an allele or stutter event and potentially result in a false exclusion at this locus.

10.4. The user specified number of contributors must be the same in both the numerator and the denominator.

10.5. Extremely trace contributors of DNA (approximately three alleles or fewer) are difficult to model. STRmix™ analysis cannot be performed on profiles containing less than 3 alleles across 2 loci.

10.6. Saturation occurs when peaks within a profile have reached the electrophoresis instrument's saturation point (i.e. an over-loaded sample). STRmix™ has a saturation setting that can account for some level of saturated data; however extreme levels of saturation are not viable for analysis. The saturation threshold for DFS laboratory's Identifiler® Plus data using the Applied Biosystems 3130xl data is 7000 rfu.

11. Documentation

11.1. STRmix™ Advanced Report(s)

12. References

12.1. STRmix™. Resolve More DNA Mixtures 2.3 User's Manual (Current Version)

12.2. Use of STRmix™ for interpretation of DNA profiles SOP, ESR Quality Documents, Version 1.0, May 18, 2015.

12.3. DFS STRmix™ Internal Validation Report, Parts I and II (2015).

12.4. Balding, D.J. and R.A Nichols, DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. Forensic Science International, (1994). 64: 125-140.

12.5. National Research Council. The Evaluation of Forensic DNA Evidence, Washington, DC: Academy Press, 1996. (colloquially referred to as “NRC II”).

12.6. SWGDAM Guidelines for the Validation of Probabilistic Genotyping Systems (Current Version)

12.7. GeneMapper®ID-X Software, User’s Manual (Current Version(s)), Human Identification Analysis, Applied Biosystems.

12.8. Applied Biosystems AmpFlSTR® Identifiler® PCR Amplification Kit User Guide (current revision).

12.9. Forensic Science Laboratory Quality Assurance Manual (Current Version)

12.10. DFS Departmental Operations Manuals (Current Versions)

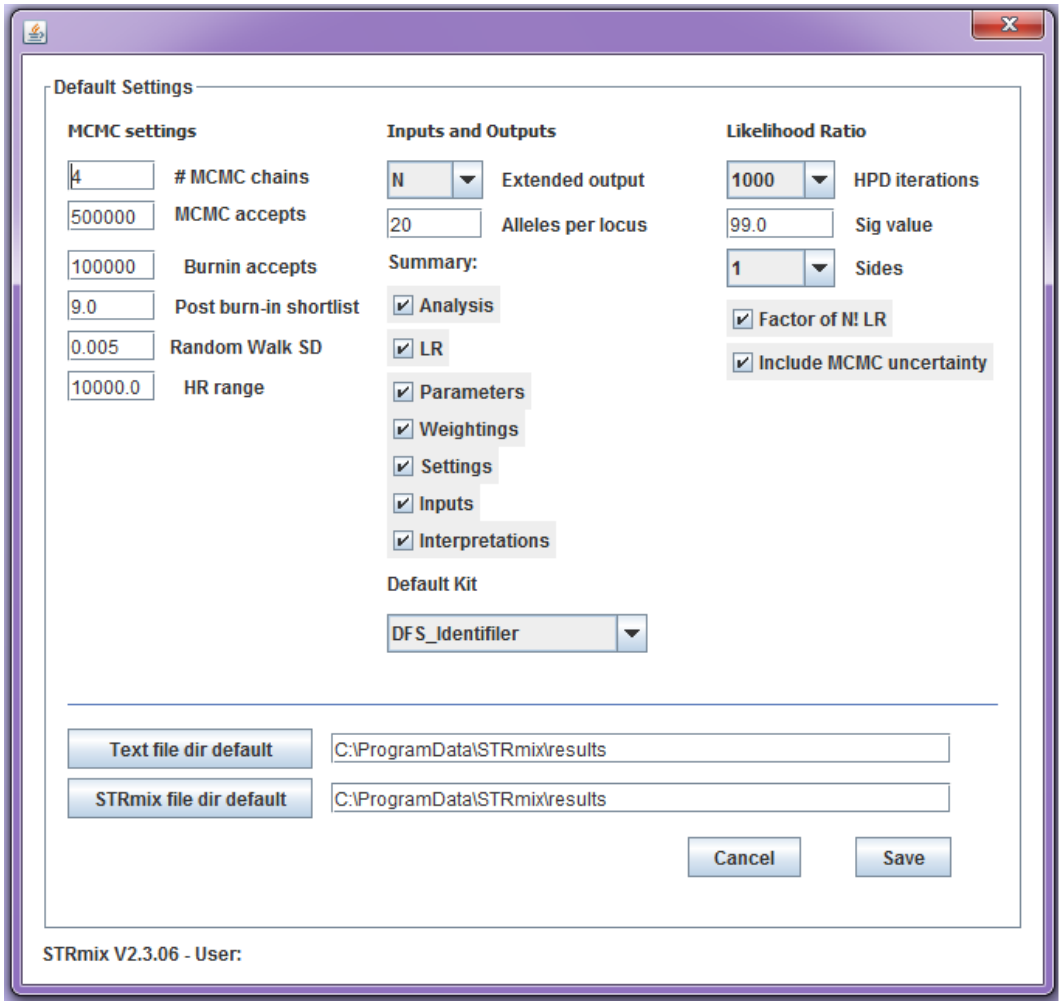
12.11. FSL Laboratory Operations Manuals (Current Versions)

12.12 FBU Quality Assurance Manual (Current Version)

13. Appendix

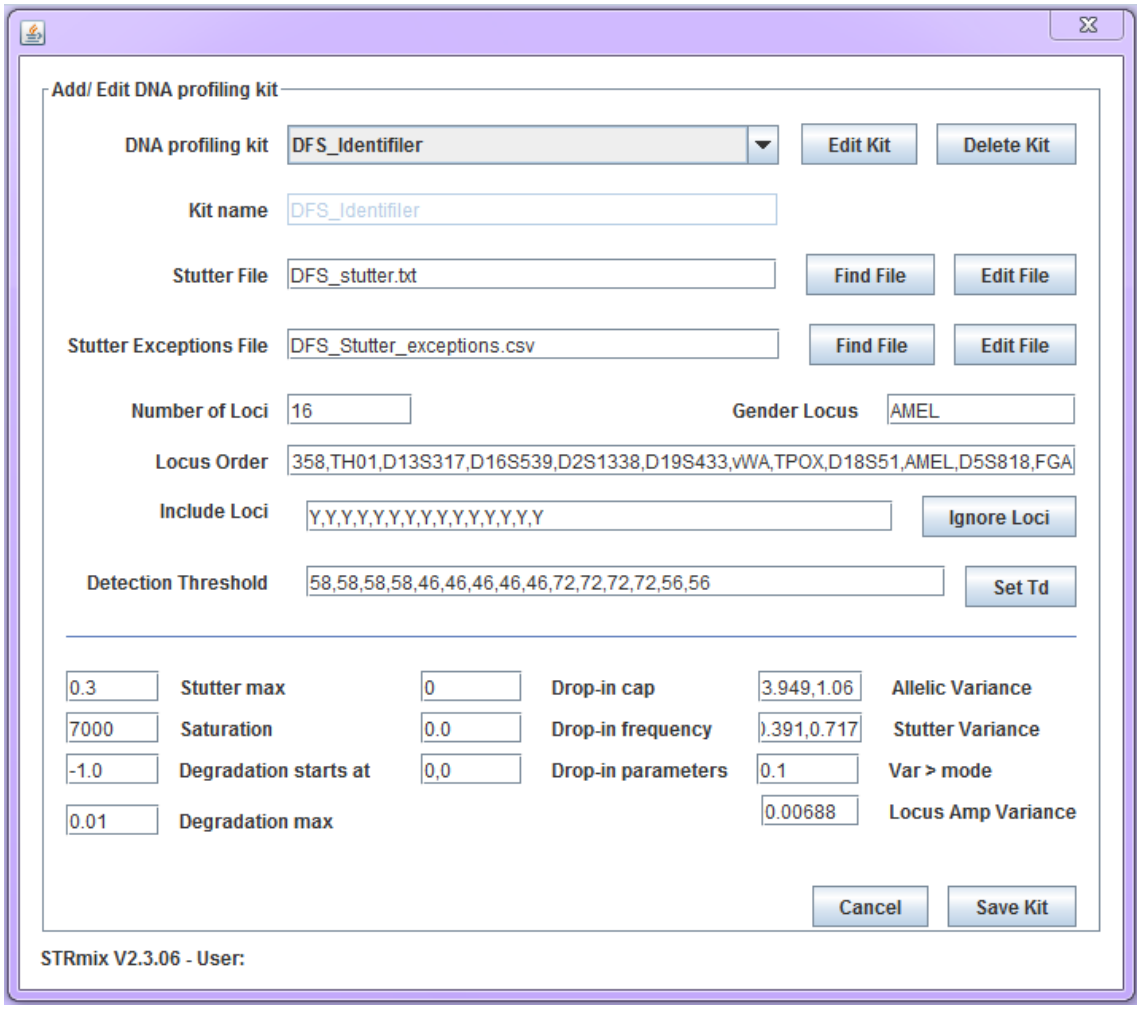
13.1. STRmix™ Default Settings:

13.1.1. **Note:** The default settings are to remain unchanged.



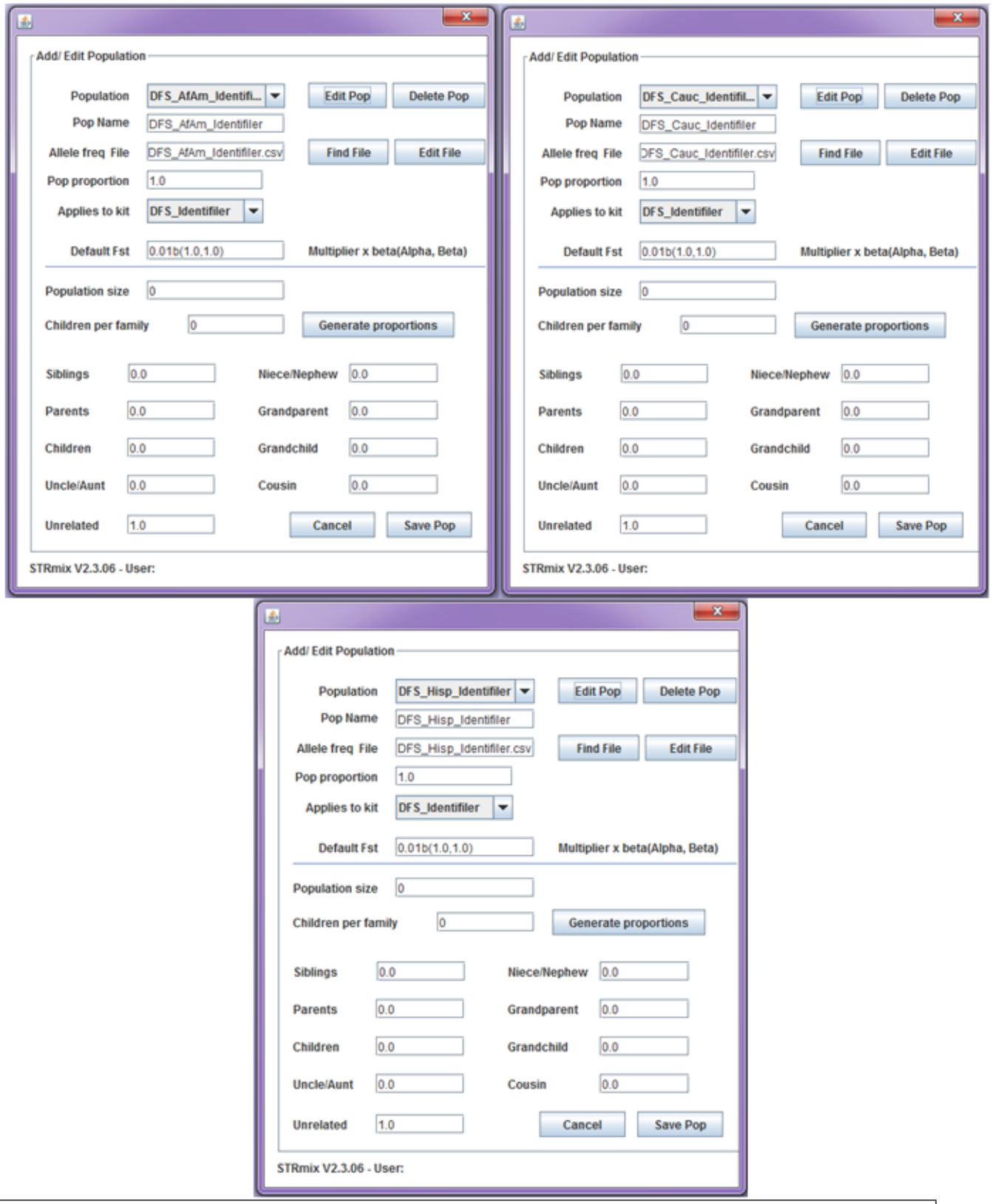
13.2. Identifiler® Settings:

13.2.1. The Identifiler® settings are to remain unchanged.



13.3. Population Data Settings:

13.3.1. The population data settings are to remain unchanged.



13.4. Allele frequency files were generated and verified using the published population data from the AmpF*l*STR® Identifiler® User's Manual (Revision D). These files can be found in the electronic documentation of STRmix™ Validation Part I: Estimation of STRmix™ Parameters.

13.5. NRC II recommendation 4.2

Balding and Nichols model also known as NRC II recommendation 4.2

Heterozygote

$$\frac{2\left[\theta+(1-\theta)p_i\right]\left[\theta+(1-\theta)p_j\right]}{(1+\theta)(1+2\theta)}$$

Homozygote

$$\frac{\left[3\theta+(1-\theta)p_i\right]\left[2\theta+(1-\theta)p_i\right]}{(1+\theta)(1+2\theta)}$$